

MULTI-DOCUMENT SUMMARIZATION SYSTEM: USING FUZZY LOGIC AND GENETIC ALGORITHM

Ansamma John

Research Scholar, Department of Computer Science

University of Kerala

Kariavattom, Kerala, India

ABSTRACT

In the recent times, the requirement for generation of multi-document summary has gained a lot of attention among the researchers. Mostly, the text summarization technique uses the sentence extraction technique where the salient sentences in the multiple documents are extracted and presented as a summary. In our proposed system, we have developed a sentence extraction based automatic multi-document summarization system that employs fuzzy logic and Genetic Algorithm (GA). At first, the different features are used to identify the significance of sentences in such a way that, each sentence in the documents is specified with the feature score. The feature score is then fed to the fuzzy logic (an AI technique) in which the fuzzy inference engine decides the importance of the sentences based on the fuzzy rules. The fuzzy rules are optimized with the help of GA algorithm and the extraction of sentences can be done based on the fuzzy score of each sentences. A multi document summary is created from the extracted sentences after removing the redundant sentences. The experiments have been done using the DUC 2002 dataset and the summary is evaluated with the measures such as Precision, Recall and F-measure.

Key words: Multi-document, Summary, Feature, Redundancy, Fuzzy logic, Fuzzy Rule, Genetic Algorithm (GA), Chromosome, Optimization, DUC 2002 Dataset (Document Understanding Conferences).

Cite this Article: Ansamma John, Multi-Document Summarization System: Using Fuzzy Logic and Genetic Algorithm. *International Journal of Advanced Research in Engineering and Technology*, 7(1), 2016, pp. 30-40.

<http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=7&IType=1>

1. INTRODUCTION

The amount of information is getting enlarged day by day, resulting in information overload. In other words, to utilize the information effectively is a challenging practical task. An urgent need for text summarization has materialized due to information overload [1]. Text summarization relates to the process of obtaining a textual document, obtaining content from it, and providing the necessary content to the user in a shortened form and in a receptive way to the requirement of user or application [2]. The technology eases the inconvenience of information overload because only a concise review has to be considered instead of a complete textual document [3]. From the early stages of text summarization, its main purpose was to assist user find information by condensing the vital information from a fundamental source and providing its shortened form. In this regard, text summarization is regarded as a mediator between the user and information included in several documents [4]. However, text summarization is still under research and has so far dealt with news text. But it is turning out to be a useful tool for information search and choosing in a diverse media.

Recently, automatic summarization has turned out to be a prominent application. This is because of the large quantity of information on the Web [8]. Automatic Text Summarization is a method in which a computer summarizes a text. A text is provided to the computer and it returns a concise and redundant-less extract of the original text. Summaries originate from two categories of text sources, a single document or a document sets [9]. Single document summarization can be defined as the process of creating a summary from a single text document. Multi-document summarization is the method of shortening, not just a single document, but a collection of related documents, into a single summary [10]. Commonly, a precise summary should be pertinent, short and articulate. In other words, the summary should meet the major concepts of the original document set, should be redundant-less and ordered [11]. These attributes are the basis of the generation process of the summary. The quality of summary is sensitive for those attributes relating to how the sentences are scored on the basis of the employed features. Consequently, the estimation of the efficacy of each attribute could result the mechanism to distinguish the attributes possessing high priority and low priority [12].

A multi-document summary possesses some notable merits over a single-document summary. It offers a domain summary of a topic based on a document set representing identical information in several documents, distinct information in separate documents, and association between sections of information across various documents. It can enable the user to look in for more information on certain facets of interest, and look into the distinctive single-document summaries [10]. Most of the similar techniques employed in single-document summarization are also employed in multi-document summarization. There exist some notable disparities [13]: (1) The degree of redundancy contained in a group of topically-related articles is considerably greater than the redundancy degree within an article, since each article is appropriate to illustrate the most important point and also the required shared background. So, anti-redundancy methods play a vital role. (2) The compression ratio (that is the summary size with regard to the size of the document set) will considerably be lesser for a vast collection topically related documents than for single document summaries. When compression demands get intensified, summarization becomes challenging. (3) The co-reference problem in summarization possesses still bigger challenges for multi-document than for single-document summarization [14].

In this paper, we have developed an automatic multi-document summarization system, which employs fuzzy logic and Genetic algorithm. Here, we have used eight different features to identify the significance of sentences in such a way that each sentence in the documents is specified with the feature score. Subsequently, the feature score is applied to the fuzzy logic, an AI technique, in which the fuzzy inference engine decides the importance of the sentences based on the fuzzy rules. The optimized rules generated by the Genetic algorithm are used as fuzzy rules. The sentences are then extracted based on the fuzzy score of the sentences and the extracted sentences make up a multi-document summary after removing the redundant sentences. We have used DUC 2002 dataset to evaluate the summarized results based on the measures such as Precision, recall and f-measure.

The rest of the paper is organized as follows: The review of related researches is given in section 2. The proposed automatic summarization system is presented in section 3. The experimental results and analysis are given in section 4. Finally, the conclusions are summarized in section 5.

2. REVIEW OF RELATED RESEARCHES

A handful of researches are available in the literature to summarize the multiple documents. Recently, several researches have been presented a multi-document summarization system based on Artificial Intelligence techniques. Some of the works presented in the multi-document summarization are given as follows:

Dragomir R. Radev *et al.* [15] have presented a multi-document summarizer, MEAD, which created summaries by employing cluster centroids generated by topic detection and tracking system. It discussed two techniques, a centroid-based summarizer, and an evaluation scheme on the grounds of sentence utility and subsumption. The assessment was subjected to single and also multiple document summaries. In the end, they elaborated about two user studies that test the models of multi-document summarization. Marie-Francine Moens *et al.* [22] have analyzed and discussed about the technologies for single and multi-document summarization which can be employed on heterogeneous texts for diverse summarization tasks. They have attributed the removal of main sentences from the documents, compressing the sentences to the appropriate content, and identifying redundant content throughout the sentences.

Fu Lee Wang *et al.* [16] have presented a multi-document summarization system to obtain the critical information from terrorism incidents. News articles of a terrorism happening were arranged into a hierarchical tree structure. Fractal summarization model was used to produce a summary for all the news stories. Experimental results proved that the system efficiently extracted the main information for the incident. Dexi Liu *et al.* [17] have proposed the multi-document summarizer employing genetic algorithm-based sentence extraction (SBGA) regards summarization process as an optimization problem where the optimal summary was selected among a summary sets created by the conjunction of the original articles sentences. To unravel the NP hard optimization problem, SBGA employed genetic algorithm, which could select the optimal summary on global aspect. To enhance the correctness of term frequency, SBGA used a TFS method, which considered word sense while determining term frequency. The experiments on DUC04 data proved that their strategy was efficient and the ROUGE-1 score was only 0.55% lesser than the best one in DUC04.

3. A SYSTEM FOR MULTI-DOCUMENT SUMMARIZATION BASED ON FUZZY LOGIC AND GENETIC ALGORITHM

Multi-document summarization is an automatic process that aims to extract the relevant summary from the multiple documents that are written about the same events. The automated procedures for generation of single document summary have been introduced in 1950's but still it has been received considerable attention among the researchers. Since the contents on the web are growing very rapidly, there is a strong need to summarize a large set of documents in a short period of time. So, several researchers [22-27] have been successfully made use of automated procedure for generating a relevant, concise and fluent summary from the multiple documents. In this research, we have developed an automated multi-document summarization system, which utilizes the fuzzy logic based summarization where the rules are optimized by the Genetic algorithm.

The steps involved in the proposed system for producing a multi-document summary from the multiple documents are,

1. Preprocessing
2. Computation of feature score using different features
3. Fuzzy modeling and generation of fuzzy rules using GA
4. Removal of Redundant Sentences

3.1. Preprocessing

The multiple documents are preprocessed using the techniques namely, Sentence segmentation, removing stop words and stemming. Thus, each sentence with their corresponding ID and the words containing in each sentence are extracted.

- *Sentence segmentation*: Separation of each sentences using the delimiter (“.” full stop).
- *Stopword Removal*: Removes of stop (linking) words like “have”, “been”, “it”, “can”, “may”, “and”, “by”, “from”, “of”, “the”, “to”, “with” and the like from the document [19].
- *Stemming algorithm*: Removes the prefixes and suffixes of each word [18].

3.2. Computation of Feature Score Using Different Features

The preprocessed documents are then utilized to compute the feature score for every sentence in accordance with the eight different features. The different features taken for our proposed system are as follows:

1. *Word similarity among documents*: A sentence is assigned by a high score based on the similar terms (words) among all the documents and the high frequency count. Here, we take the top n -frequent words from every documents and identify the similar keywords in the frequent word list. For every sentence, we count the number of occurrences of these similar identified keywords. Feature score, word similarity among document is calculated by the ratio of the similar keywords count of the given sentences to the number of frequent words (n) is taken to find the similarity.
2. *Centroid value*: Centroid [15] is a feature value which is used to identify the salient sentences for summarizing the multiple documents. The centroid value for each sentence (C_s) is the sum of the individual centroid of the words (C_w) containing in the sentence. The centroid of each word or term is the product value of term frequency (TF) and the inverse document frequency (IDF). The

term frequency (TF) is the number of occurrences of a given term appeared in the document. The inverse document frequency (IDF) is obtained by the division of the number of documents in the document set and the number of documents containing the given word, and then find the logarithm of that quotient.

$$C_s = \sum_{i=1}^m C_{w_i}$$

$$C_w = TF * IDF$$

$$IDF = \log \left[\frac{\text{Number of documents}}{\text{Number of documents containing the given word}} \right]$$

Where, $C_s \rightarrow$ Centroid value of the sentence

$C_w \rightarrow$ Centroid value of the word

$TF \rightarrow$ Term Frequency

$IDF \rightarrow$ Inverse Document Frequency

3. *Paragraph Frequency:* The words are extracted from individual paragraphs and also identify the number of occurrences of the extracted words among the paragraphs. The number of occurrences of keywords is summed up to get the paragraph frequency of the given paragraph. The feature score is computed by the ratio of the paragraph frequency of the given paragraph to the maximum paragraph frequency in the document. The feature score for each sentence in the given paragraph is same as the feature score of the given paragraph.
4. *Positional based score:* The maximized centroid value is given as a score value for the first sentence in the given document and the remaining sentences get the score value based on their corresponding position in the document. The positional score P_{s_k} for all sentences within a document calculated is given in [15]:

$$P_{s_k} = \frac{n-k+1}{n} * C_{\max}$$

Where, $C_{\max} \rightarrow$ Maximized centroid value

$n \rightarrow$ Number of sentences in a document

5. *Format based score:* In general, the important words of the sentences are represented with the specific formats like Italics, Bold, underlined and different font sizes. The words represented with the above specific formats are likely important for the summarized result. The feature score is the ratio of the number of words in the sentence with special format to the total number of words in the sentence.
6. *Numerical data:* The numerical data presented in the document have some significant information and it would more likely include in the summary. The sentences contained numerical data are the most preferable sentences. The feature score is calculated by the ratio of the number of numerical data occurred in the sentence and the length of the sentence.

7. *Title features:* The title words are probably an important feature when summarizing the document. The feature score for this feature is computed as the ratio of the number of similar title words in the sentence to the total number of words in the title.
8. *Sentence length:* This feature is useful for removing the short sentences. Generally, the short sentences are the news article contains datelines and author name. But, it is not necessary for the summary. So, the sentence length feature has given more priority for the long length sentences rather than the short one. It is defined as the ratio of number of words occurring in the sentences and the number of words in the longest sentence in the document.

3.3. Fuzzy Modeling and Generation of Fuzzy Rules Using GA

The term "fuzzy logic" resulted in the development of the theory of fuzzy sets by Zadeh [20]. The fuzzy logic is extension of the classical logic in form of generalization of the classical logic inference rules which has capability to deal with approximate reasoning. The fuzzy set is an expansion for the traditional set "crisp set" in which each member has a degree of membership to that set determined by membership function. The membership function is a function that gives membership degree to each member in the target set, the range of membership degree between zero and one. The fuzzy logic has benefit in terms of simplicity of development and modification because the rules are well understandable and simple to modify, add novel rules or remove existing rules [21].

Here, we make use of the fuzzy logic system in the proposed multi-document summarization system. The sentences presented in the documents are assigned by a feature score using the aforementioned eight features. The feature score of each sentence are given as an input to the fuzzy logic system. The advantages of using fuzzy logic system are (1) Allows imprecise/contradictory inputs, (2) Permits fuzzy thresholds, (3) Reconciles conflicting objectives (4) Rule base or fuzzy sets easily modified. The fuzzy logic system consists of four components: (a) Fuzzifier (b) Rule base (c) inference engine (c) Defuzzifier.

3.3.1. Fuzzifier

The obtained feature score of every sentence is fed to the fuzzifier, which converts the numerical data into the linguistic values (High, medium, low). The linguistic values of each feature score is obtained using the membership function, which is a curve that defines how each point in the input feature score is converted into a membership value (or degree of membership) between 0 and 1. Here, we use the triangle membership function which is defined as follows,

$$f(x:a,b,c) = \begin{cases} 0 & , \text{ if } x < a, x > c \\ \frac{x-a}{b-a} & , \text{ if } a \leq x \leq b \\ \frac{a-x}{b-a} & , \text{ if } b \leq x \leq c \end{cases}$$

Where a, b and c are characteristic parameters of a fuzzy set.

3.3.2 Fuzzy Rule Base

Once the inputs have been fuzzified, we define the fuzzy rules which are important for any fuzzy system. In general, fuzzy logic based summarization system has used

the manually defined fuzzy rules. This approach is not an effective one when the number features and linguistic variable is large. In order to obtain the effective rules, we have used Genetic algorithm to provide optimized rules for the fuzzy system. The obtained optimized rules are then stored in the fuzzy rule base.

3.3.3 Generation of Fuzzy Rules Using Genetic Algorithm

Genetic algorithm is used in various applications to find the optimal solution. Genetic algorithm (GA) is an evolutionary algorithm that evolves computer programs and predicts mathematical models from experimental data. GA starts with a random population of candidate solutions in the form of chromosomes. The chromosomes are then evaluated based on a fitness value and chosen by fitness to reproduce with modification via genetic operations such as crossover and mutation. The new generation of solutions goes through the same process until the termination criteria is satisfied. The fittest individual serves as the final solution. We have utilized the GA algorithm to find the optimal rules for fuzzy system to summarize the multiple documents.

- *Chromosome initialization:* Initially, the set of chromosomes are generated randomly and each chromosome consists of eight genes that represent the linguistic variable of each features.
- *Fitness function:* Fitness function is used to evaluate the survival of the chromosomes and the fitness function is computed as follows,

$$F_T = \frac{\sum_{J \in H, L, M} G_m(J)}{\sum_{J \in H, L, M} G(J)}; J = \{\text{High, Medium, Low}\}$$

Where, $F_T \rightarrow$ Fitness function

$G_m(J) \rightarrow$ Number of genes co-occurring in chosen chromosome and the reference chromosome

$G(J) \rightarrow$ Number of genes presented in the reference chromosome.

- *Crossover and Mutation:* In the first level of iteration, two chromosomes are generated randomly and the fitness value is calculated for the randomly generated two chromosomes. The crossover and mutation operation is applied over the randomly generated chromosomes. Here, we used the single point cross over and mutation operation. Then, the fitness value is computed for the newly generated chromosomes and the better chromosome is selected from the first iteration. The better chromosome from the first iteration is to be given to the next level of iteration with one randomly generated chromosome. Again, crossover and mutation is done with the above two chromosomes and the better one is selected based on the fitness value. This process is repeated for ‘ n ’ number of iterations.
- *(d) Chromosome selection:* We obtained the better set of chromosomes once the ‘ n ’ number of iteration gets terminated and they are sorted based on the fitness value. The optimized chromosomes are taken from the sorted list of chromosomes. Each optimized chromosome represents the fuzzy rule, which is used for finding the importance of sentences.

3.3.4. Inference Engine

The inference engine is used to provide the score for every sentence based on the fuzzy rules enclosed in the fuzzy rule base. It provides the fuzzy set {important, unimportant}, which means whether the sentences are important or unimportant. The fuzzy rule enclosed in the rule base consists of two parts antecedent and consequent. Antecedent is defined by the multivariate membership functions and consequent is the inference of the rule which determines whether the sentence is important or unimportant based on the input. The fuzzy rule is expressed as: IF (Word similarity among documents is H) and (Centroid value is H) and (Positional based score is H) and (Paragraph Frequency is M) and (Format based score is L) and (Numerical data is H) and (Title features is M) and (Sentence length is L) THEN (Sentence is important).

3.3.5. Defuzzifier

The input for the defuzzifier is a fuzzy set {important, unimportant} and the output is a crisp value. The resultant crisp value is the final score of each sentence.

3.4 REMOVAL OF REDUNDANT SENTENCES

The sentences that are extracted from the multiple documents based on the final score generated by defuzzifier. The extracted sentences constitute the summary which may consist of redundant sentences, due to the fact that the multiple documents are written about the same events. So it is necessary that the redundant sentences must be removed from the summary. The redundant sentences are identified by the similar words occurring among the sentences. The formula used to perform the redundant removal process in between two sentences is given in [15],

$$R = \frac{2 * N_s}{N_1 + N_2}$$

Where, $R \rightarrow$ Redundancy of two sentences

$N_s \rightarrow$ Number of similar words in the first and second sentences

$N_1 \rightarrow$ Number of words in the first sentence

$N_2 \rightarrow$ Number of words in the second sentence

4. EXPERIMENTAL RESULTS AND ANALYSIS

This section describes the experimental results of the proposed multi-document summarization system. The proposed system is implemented in MATLAB (Matlab7.8). We have used the DUC 2002 dataset in our experiments to generate the multi-document summary.

4.1 EVALUATION MEASURE

The proposed multi-document summarization system has utilized the evaluation measures such as precision, recall and F-measure for validating our proposed approach. For evaluation, we have used the summary presented in the dataset. (1) *Precision*: It is the ratio of the number of similar sentences in both the summaries and the number of sentences in the summary generated by the proposed system. (2) *Recall*: It is the ratio of the number of similar sentences in both the summaries and the number of sentences in the summary presented in the dataset. (3) *F-measure*: F-

measure is a measure that combines Precision and Recall. It is defined by the following equation.

$$F_m = 2 * \frac{P * R}{P + R}$$

Where, $F_m \rightarrow$ F-measure

$P \rightarrow$ Precision

$R \rightarrow$ Recall

4.2 Experimental Results

The experimental results of the proposed multi-document summarization system are presented in this section. The multi-documents (written about the same topic) available in the DUC 2002 dataset is used in our proposed multi-document summarization system that generates the multi-document summary. The generated summary is evaluated with the summary presented in the dataset using the measures such as precision, recall and the F-measure. The evaluation measures are computed for different percentage in summary (with respects to documents size) and the results are given in table 1. The corresponding graph is shown in figure 2.

Table 1 Evaluation measures for % in summary

% in summary	Precision	Recall	F-measure
40	0.565	0.43	0.488
50	0.517	0.5	0.508
60	0.5	0.533	0.515

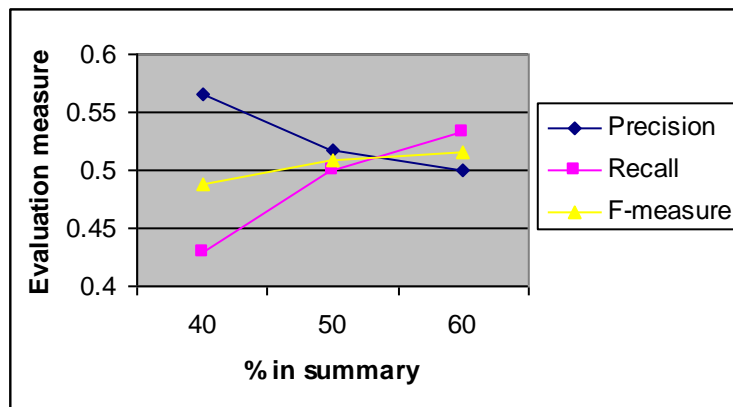


Figure 2 Evaluation measures vs. % in summary

5. CONCLUSION

We have developed automatic multi-document summarization system which incorporates the fuzzy logic and GA algorithm. We have used eight different features for feature extraction phase. The feature score of the sentences is applied to the fuzzy logic system in which the fuzzy rules are optimized with the help of Genetic Algorithm. We have used DUC 2002 dataset to evaluate the summarized results based

on the measures such as Precision, recall and f-measure. The experiment results showed that the incorporation of fuzzy logic with GA effectively summarize the multi-documents.

REFERENCES

- [1] Barzilay. R., Elhadad, N and McKeown. K. R., "Inferring strategies for sentence ordering in multidocument news summarization", Journal of Artificial Intelligence Research, vol.17, pp. 35–55, 2002.
- [2] Inderjeet Mani, "Automatic summarization", Kluwer Academic Publishers, 2004.
- [3] Jen-Yuan Yeh , Hao-Ren K and Wei-Pang Yang , "iSpreadRank: Ranking sentences for extraction-based summarization using feature weight propagation in the sentence similarity network", Expert Systems with Applications, vol. 35, no. 3, pp. 1451-1462, October 2008.
- [4] Yatsko V., Shilov S. and Vishniakov T., "A Semi-automatic Text Summarization System", In proceedings of the 10th International Conference on Speech and Computer, Patras, pp. 283-288, 2005.
- [5] Kaustubh Patil and Pavel Brazdil, "SUMGRAPH: Text Summarization Using Centrality In The Pathfinder Network", International Journal on Computer Science and Information Systems, vol.2, no.1, pp. 18-32, 2007.
- [6] Ladda Suanmali, Naomie Salim and Mohammed Salem Binwahlan, "Fuzzy Logic Based Method for Improving Text Summarization", International Journal of Computer Science and Information Security (IJCSIS), vol. 2, no. 1, 2009.
- [7] Ladda Suanmali, Naomie Salim and Mohammed Salem Binwahlan, "Automatic Text Summarization Using Feature Based Fuzzy Extraction", vol. 20, no. 2, pp. 105-115, November 2009.
- [8] F. Canan Pembe and Tunga Gungor, "Automated Query-biased and Structure-preserving Text Summarization on Web Documents", in proceedings of the International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, June 2007.
- [9] Liang Zhou, Miruna Ticea and Eduard Hovy, "Multi-document Biography Summarization", in Proceedings of Empirical Methods in Natural Language Processing, 2004.
- [10] Shiyen Ou, Christopher S.G. Khoo and Dion H. Goh, "Design and development of a concept-based multidocument summarization system for research abstracts", Journal of Information Science, vol. 34 , no. 3, pp. 308-326 , June 2008.
- [11] You Ouyang, Wenji Li and Qin Lu, "An Integrated Multi-document Summarization Approach based on Word Hierarchical Representation", in proceedings of the ACL-IJCNLP, singapore, pp. 109–112, 2009.
- [12] Mohammed Salem Binwahlan, Naomie Salim and Ladda Suanmali, "Swarm Based Features Selection for Text Summarization", IJCSNS International Journal of Computer Science and Network Security, vol. 9, no.1, January 2009.
- [13] Jade Goldstein, Vibhu Mittal, Jaime Carbonell and Mark Kantrowitz, "Multi-Document Summarization by Sentence Extraction", NAACL-ANLP 2000 Workshop on Automatic summarization, Seattle, Washington, vol. 4, pp. 40 - 48, 2000.

- [14] Breck Baldwin and Thomas S. Morton, "Dynamic coreference-based summarization", in Proceedings of the Third Conference on Empirical Methods in Natural Language Processing (EMNLP-3), Granada, Spain, June 1998.
- [15] Dragomir R. Radev a, Hongyan Jing, Magorzata Sty and Daniel Tam, "Centroid-based summarization of multiple documents, Information Processing and Management, vol. 40, no.6, pp. 919–938, 2004.
- [16] Fu Lee Wang, Christopher C. Yang and Xiaodong Shi," Multi-document Summarization for Terrorism Information Extraction", Lecture Notes in Computer Science", Springer, vol. 3975, May 2006.
- [17] Dexi Liu, Yanxiang He, Donghong Jiand Hua Yang, "Genetic Algorithm Based Multi-document Summarization", Lecture Notes in Computer Science", Springer, vol. 4099, pp. 1140-1144, July 2006.
- [18] Lovins, J.B. 1968: "Development of a stemming algorithm". Mechanical Translation and Computational Linguistics, vol. 11, pp. 22-31, 1968.
- [19] Pant, G., Srinivasan, P., Menczer, F., "Crawling the Web", Web Dynamics: Adapting to Change in Content, Size, Topology and Use, edited by M. Levene and A. Poulouvassilis, Springer- verilog, pp. 153-178, November 2004.
- [20] Zadeh, L.A., "Fuzzy sets", Information and control, vol.8, pp. 338-353, 1965.
- [21] Klir, G.J. and B. Yuan, "Fuzzy Sets and Fuzzy Logic: Theory and Applications", 1st Edn., Prentice-Hall, New York, pp: 592, 1995.
- [22] Marie-Francine Moens,Roxana Angheluta and Jos Dumortier, "Generic technologies for single- and multi-document summarization", Information Processing and Management, vol. 41, no.3, pp.569-586, May 2005.
- [23] Jiaming Zhanb, Han Tong Lohb and Ying Liua, "On macro- and micro-level information in multiple documents and its influence on summarization", International Journal of Information Management, Vol. 29, No. 1, pp. 57-66, February 2009.
- [24] Sugandhi Midha, Comparative Study of Remote Sensing Data Based on Genetic Algorithm. *International Journal of Computer Engineering and Technology*, **5**(1), 2014, pp. 141-152.
- [25] Palak Sharma, Shelza, Gurjit Singh Randhawa and Rajinder Singh Virk, Cost Optimization of Pipeline Systems Using Genetic Algorithm, *International Journal of Computer Engineering and Technology*, **4**(4), 2013, pp. 485-494.
- [26] David M. Zajica, Bonnie J. Dorra and Jimmy Linb, "Single-document and multi-document summarization techniques for email threads using sentence compression", Information Processing and Management, Vol. 44, No. 4, pp. 1600-1610, July 2008.
- [27] Lin Zhao, Lide Wua, and Xuanjing Huang, "Using query expansion in graph-based approach for query-focused multi-document summarization", Information Processing and Management, Vol. 45, No. 1, pp. 35-41, January 2009.
- [28] Yohei Seki, Koji Eguchi and Noriko Kando, " Multi-Document Viewpoint Summarization Focused on Facts, Opinion and Knowledge", Computing Attitude and Affect in Text: Theory and Applications, Springer, Vol. 20, pp.317-336, January 2006.
- [29] Florian Boudin and Juan Manuel Torres Moreno, " NEO-CORTEX: A Performant User-Oriented Multi-Document Summarization System", Lecture Notes in Computer Science", Springer, Vol 4394, pp.551-562, May 2007.